

Causal network inference using biochemical kinetics

C. J. Oates^{1,*}, F. Dondelinger², N. Bayani³, J. Korkola⁴, J. W. Gray⁴ and S. Mukherjee^{2,*}

¹Department of Statistics, University of Warwick, Coventry, UK.

²MRC Biostatistics Unit and CRUK Cambridge Institute, University of Cambridge, Cambridge, UK.

³Lawrence Berkeley National Laboratory, University of California, Berkeley, USA.

⁴Knight Cancer Institute, Oregon Health and Science University, Portland, USA.

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXXX

ABSTRACT

Motivation: Network models are widely used as structural summaries of biochemical systems. Statistical estimation of networks is usually based on linear or discrete models. However, the dynamics of these systems are generally nonlinear, suggesting that suitable nonlinear formulations may offer gains with respect to network inference and associated prediction problems.

Results: We present a general framework for both network inference and dynamical prediction that is rooted in nonlinear biochemical kinetics. This is done by considering a dynamical system based on a chemical reaction graph and associated kinetics parameters. Inference regarding both parameters and the reaction graph itself is carried out within a fully Bayesian framework. Prediction of dynamical behavior is achieved by averaging over both parameters and reaction graphs, allowing prediction even when the underlying reactions themselves are unknown or uncertain. Results, based on (i) data simulated from a mechanistic model of mitogen-activated protein kinase signaling and (ii) phosphoproteomic data from cancer cell lines, demonstrate that nonlinear formulations can yield gains in network inference and permit dynamical prediction in the challenging setting where the reaction graph is unknown.

Availability: MATLAB R2014a software is available to download from warwick.ac.uk/chrisoates.

Contact: c.oates@warwick.ac.uk; sach@mrc-bsu.cam.ac.uk

1 INTRODUCTION

Statistical network inference techniques are widely used in the analysis of multivariate biochemical data (Sachs *et al.*, 2005; Ellis and Wong, 2008). These techniques aim to make inferences regarding a network N whose vertices are identified with biomolecular components (e.g. genes or proteins) and edges with (direct or indirect) regulatory interplay between those components.

Algorithms for network inference are typically rooted in linear or discrete models whose statistical and computational advantages facilitate exploration of large spaces of networks (Werhli *et al.*, 2006). On the other hand, when the network topology is known, nonlinear ordinary differential equations (ODEs) are widely used to model biochemical dynamics (Kholodenko, 2006; Chen *et al.*,

2009). The intermediate case where biochemical ODE models are employed to select between network models has received less attention.

We propose a general framework called “Chemical Model Averaging” (CheMA) that uses biochemical ODE models to carry out both network inference and dynamical prediction. In summary, we consider a dynamical system $d\mathbf{X}/dt = \mathbf{f}_G(\mathbf{X}, \boldsymbol{\theta})$ where the state vector \mathbf{X} contains the abundances of molecular species, G is a chemical reaction graph (Craciun and Pantea, 2008) that characterizes reactions in the system, \mathbf{f}_G is a kinetic model that depends on G and $\boldsymbol{\theta}$ collects together all unknown kinetic parameters. A network N is obtained as a coarse summary $N(G)$ of the reaction graph G in which each chemical species appears as a single node and directed edges indicate that the parent is involved in chemical reaction(s) which have the child as product (we make these notions precise below). Given time-course data \mathcal{D} consisting of noisy measurements of \mathbf{X} , we carry out inference and prediction within a fully Bayesian framework. Information on the kinetic parameters is integrated through a prior density $p(\boldsymbol{\theta}|G)$. In particular we treat G itself as unknown and make inference concerning it using the posterior distribution,

$$p(G|\mathcal{D}) \propto p(G) \underbrace{\int p(\mathcal{D}|\boldsymbol{\theta}, G) p(\boldsymbol{\theta}|G) d\boldsymbol{\theta}}_{\text{marginal likelihood } p(\mathcal{D}|G)} \quad (1)$$

where the marginal likelihood $p(\mathcal{D}|G)$ captures how well the chemical reaction graph G describes data \mathcal{D} , taking into account both parameter uncertainty and model complexity. In contrast to linear or discrete models that are motivated by tractability, our likelihood $p(\mathcal{D}|\boldsymbol{\theta}, G)$ depends on (richer) reaction graphs G and their associated kinetics.

This paper makes three contributions: (1) A general framework for joint network learning and dynamical prediction using ODE models; (2) a specific implementation (CheMA 1.0), rooted in Michaelis-Menten kinetics, that uses Metropolis-within-Gibbs sampling to allow fully Bayesian inference at feasible computational cost; and (3) an empirical investigation, using both simulated and experimental time-course data, of the performance of CheMA 1.0 relative to several existing approaches for network inference and dynamical prediction.

*to whom correspondence should be addressed

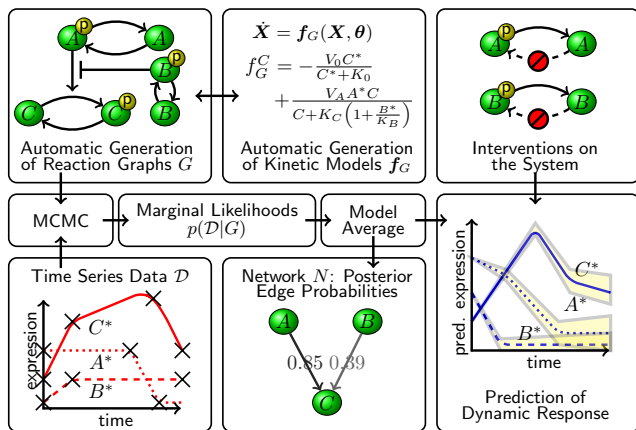


Fig. 1: Chemical Model Averaging (CheMA). Chemical reaction graphs G summarize interplay that is described quantitatively by kinetic equations f_G . Candidate graphs G are scored against observed time course data \mathcal{D} in a fully Bayesian framework. A network N gives a coarse summary of the system; marginal posterior probabilities of edges in N quantify evidence in favor of causal relationships. The reaction graph G (and N) is treated as an unknown, latent object and the methodology allows fully Bayesian prediction of dynamics (including under intervention) in the unknown graph setting.

The statistical connection between *linear* ODEs and network inference using linear models has been discussed in Oates and Mukherjee (2012) and exploited in Gardner *et al.* (2003); Bansal *et al.* (2006). Several approaches based on *nonlinear* ODEs have been proposed, including Nachman *et al.* (2004); Nelander *et al.* (2008); Åijö and Lähdesmäki (2010); Honkela *et al.* (2010). Our contribution differs to these by formulating a fully Bayesian approach to both network inference and dynamical prediction that is rooted in chemical kinetics. Bayesian model selection based on nonlinear ODE has been shown to be a promising strategy for elucidation of specific signaling mechanisms (e.g. Xu *et al.*, 2010). The work we present differs in motivation and approach in that we exploit automatically-generated rather than hand-crafted biochemical models, thereby allowing full network inference without manual specification of candidate models. Oates *et al.* (2012) performed Bayesian model selection by comparing steady-state data to equilibrium solutions of automatically generated ODE models. This paper extends this approach to time course data and prediction of dynamics.

There are several considerations that motivate CheMA: (i) Inference in biological systems is complicated by strong correlations between components that are co-regulated but not causally linked. It is well known that, under a linear formulation, the causal network N is generally unidentifiable (Pearl, 2009). For example, it may not be possible to orient certain edges, or edges may be inferred between co-regulated nodes due to strong associations between them. Nonlinear kinetic equations, in contrast, are able to confer asymmetries between nodes and may be sufficient to enable orientation of all edges (Peters *et al.*, 2011), although we note that nonlinear models still require many assumptions to

allow causal inference (Pearl, 2009). As a consequence, CheMA can in principle aid in causal network inference, and empirical results below support this. (ii) In contrast to linear models, in CheMA the mechanistic roles of individual variables are respected. This facilitates analysis of data obtained under specific molecular interventions and enhances scientific interpretability. (iii) Prediction of dynamical behavior (e.g. response to a stimulus or to a drug treatment) in general depends on the chemical reaction graph. In settings where the graph itself is unknown or uncertain (e.g. due to genetic or epigenetic context), CheMA allows ensemble-averaged prediction of dynamics.

The CheMA framework is general and can in principle be used in many settings where kinetic formulations are available to describe the dynamics, including gene regulation, metabolism and protein signaling. For definiteness, in this paper we focus on protein signaling networks mediated by phosphorylation and provide a specific implementation of the general framework. Phosphorylation kinetics have been widely studied (Kholodenko, 2006) and ODE formulations are available, including those based on Michaelis-Menten kinetics (Leskovic, 2003).

The remainder of the paper is organized as follows. First, we introduce the model and associated statistical formulation. Second, we discuss network inference and dynamical prediction within this framework. Third, we show empirical results, comparing CheMA 1.0 to several existing approaches, using (i) data simulated from a mechanistic model of mitogen-activated protein kinase (MAPK) signaling, and (ii) phosphoproteomic time course data from human cancer cell lines. Finally, we discuss our findings and suggest several directions for further work.

2 METHODS

Below we describe an attempt to implement the CheMA framework, called CheMA 1.0, for the specific context of protein phosphorylation networks. Fig. 1 provides an outline of our workflow below.

2.1 Automatic Generation of Reaction Graphs G

We construct reaction graphs for p proteins $\{X_1 \dots X_p\} = \mathcal{V}$. Each X_i can be phosphorylated to X_i^* ; the set of phosphorylated proteins is \mathcal{V}^* . Phosphorylation reactions $X_i \rightarrow X_i^*$ are catalyzed by enzymes $E \in \mathcal{E}_i$; the subscript indicates that each protein may have a specific set of enzymes (kinases). We consider the case in which the kinases themselves are phosphorylated proteins, i.e. $\mathcal{E}_i \subseteq \mathcal{V}^*$ (if phosphorylation of X_i is not driven by an enzyme in \mathcal{V}^* , we set $\mathcal{E}_i = \emptyset$). For simplicity we do not consider multiple phosphorylation sites, other post-translational modifications (e.g. ubiquitinylation), protein degradation, nor spatial effects. The ability of enzyme $E \in \mathcal{E}_i$ to catalyze phosphorylation of X_i may be inhibited by proteins $I \in \mathcal{I}_{i,E} \subseteq \mathcal{V}^*$; the double subscript indicates that inhibition is specific to both target X_i and enzyme E (see below).

The reaction graph G provides a visual representation of the sets \mathcal{E}_i and $\mathcal{I}_{i,E}$; Fig. 1 shows an illustrative example on three proteins A, B and C. A causal biological network $N(G)$ is formed by drawing exactly p vertices and edges (i, j) indicating that X_i^* is either an enzyme catalyzing phosphorylation of X_j , or an inhibitor of such an enzyme. That is, $(i, j) \in N \iff i \in \mathcal{E}_j \vee \exists E \cdot i \in \mathcal{I}_{j,E}$. For the example shown in Fig. 1, the corresponding network N is the directed graph $A \rightarrow C \leftarrow B$.

2.2 Automatic Generation of Kinetic Models f_G

The reaction graph G can be decomposed into local graphs G_i describing enzymes (and their inhibitors) for phosphorylation of protein X_i . For

simplicity of exposition we consider inference concerning G_i . Thus, X_i plays the role of the substrate; following conventional notation in enzyme kinetics, we refer to X_i using the symbol S .

We use kinetic models f_G based on Michaelis-Menten functionals (Leskovic, 2003). Here we restrict attention to a relatively simple model class, but more complex dynamics could be incorporated if appropriate. The rate of phosphorylation due to kinase E is given by $V_E[E][S]^h/([S]^h + K_E^h)$, which acknowledges variation of kinase concentration $[E]$ and permits kinase-specific response profiles, parameterized by K_E and h , with rate constant V_E . In subsequent experiments the Hill coefficient h is taken equal to 1 (non-cooperative binding). We entertain competitive inhibition, where substrate and inhibitor I compete for the same binding site on the enzyme ($EI \rightleftharpoons E \rightleftharpoons ES \rightarrow E + S^*$). When multiple inhibitors are present, they are assumed to act exclusively, competing for the same binding site on the enzyme ($EI \rightleftharpoons E \rightleftharpoons EI'$), corresponding mathematically to a rescaling of the Michaelis-Menten parameter $K_E \mapsto K_E(1 + \sum_{I \in \mathcal{I}_{S,E}} [I]/K_I)$. We do not model phosphatase specificity; in particular, dephosphorylation is assumed to occur at a rate $V_0[S^*]/([S^*] + K_0)$, depending on a Michaelis-Menten parameter K_0 and taking a maximal value V_0 .

Combining these assumptions produces a kinetic model for phosphorylation of substrate S , given by $f_{G,S}(\mathbf{X}, \boldsymbol{\theta}_S) =$

$$-\frac{V_0[S^*]}{[S^*] + K_0} + \sum_{E \in \mathcal{E}_S} \frac{V_E[E][S]}{[S] + K_E \left(1 + \sum_{I \in \mathcal{I}_{S,E}} \frac{[I]}{K_I}\right)} \quad (2)$$

where the parameter vector $\boldsymbol{\theta}_S$ contains the maximum rates \mathbf{V} and Michaelis-Menten constants \mathbf{K} and the (local) graph G_S specifies the sets \mathcal{E}_S and $\mathcal{I}_{S,E}$. The complete dynamical system f_G is given by taking, for each species $S \in \mathcal{V}$, a model akin to Eqn. 2. In this way we are able to automate the generation of candidate parametric ODE models.

2.3 Model Averaging

Evidence for a causal influence of protein i on protein j is summarized by the marginal posterior probability of a directed edge (i, j) in the network N . This is obtained by averaging over all possible reaction graphs G , as

$$p((i, j) \in N | \mathcal{D}) = \frac{\sum_{G: i \in G_j} p(\mathcal{D} | G) p(G)}{\sum_G p(\mathcal{D} | G) p(G)}. \quad (3)$$

We note that whilst Eqn. 3 is an intuitive summary statistic, the full posterior over reaction graphs G is also available for more detailed exploration. In the same vein, model averaging is used to compute posterior predictive distributions (see Supplementary Information).

Following work in structural inference for graphical models (Ellis and Wong, 2008) we bound graph in-degree; in particular, we bound the number of kinases $|\mathcal{E}_S| \leq c_1$ and the number of inhibitors $|\mathcal{I}_{S,E}| \leq c_2$ (see Section 3.1 below).

Bayesian variable selection requires multiplicity correction to control the false discovery rate and avoid degeneracy (Scott and Berger, 2010). For phosphorylation networks we achieve multiplicity correction using a prior $p(G)$ uniform over the number of kinases, and for a given kinase, uniform over the number of kinase inhibitors:

$$p(G) = \prod_{i=1}^p \binom{p}{|\mathcal{E}_i|}^{-1} \prod_{E \in \mathcal{E}_i} \binom{p}{|\mathcal{I}_{i,E}|}^{-1} \quad (4)$$

We note that the above prior does not include biological knowledge concerning specific edges; informative structural priors that incorporate biological knowledge are also available in the literature (Mukherjee and Speed, 2008).

2.4 Statistical Formulation: CheMA 1.0

The CheMA framework demands a heavily computational approach to inference. Below we describe an approximate methodology, CheMA 1.0, that aims to accurately approximate the posterior expectations that are of interest.

2.4.1 Likelihood Data \mathcal{D} comprise measurements $y_i(t_j)$ and $y_i^*(t_j)$ proportional to the concentrations of unphosphorylated and phosphorylated forms, respectively, of protein i at discrete times t_j , $0 \leq j \leq n$. Data are scale-normalized to give unit mean for each protein ($\sum_j y_i(t_j) = \sum_j y_i^*(t_j) = n+1$). In CheMA 1.0, observables are related to dynamics via “gradient-matching”. We follow Bansal *et al.* (2006); Åijö and Lähdesmäki (2010); Oates and Mukherjee (2012) and employ a simple Euler scheme that approximates the gradient dX_i/dt at time t_j by $z_i(t_j) = (y_i^*(t_j) - y_i^*(t_{j-1})) / (t_j - t_{j-1})$. We note that more accurate approximations could be used, at the cost of requiring more data points or additional modelling assumptions (see Discussion). The ODE model $f_{G,S}$ (Eqn. (2)) is formulated as a statistical model by constructing, conditional upon (unknown) Michaelis-Menten parameters \mathbf{K} , a design matrix $\mathbf{D}_{G,S}(\mathbf{K})$ with rows

$$\left[-\frac{y_S^*}{y_S^* + K_0}, \dots, \underbrace{\frac{y_S^* y_S}{y_S + K_E \left(1 + \sum_{I \in \mathcal{I}_{S,E}} \frac{y_I^*}{K_I}\right)}}_{E \in \mathcal{E}_S}, \dots \right] \quad (5)$$

and then interpreting Eqn. (2) statistically as

$$\mathbf{z}_S = \mathbf{D}_{G,S}(\mathbf{K})\mathbf{V} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}) \quad (6)$$

where $\mathbf{z}_S = [z_S(t_1), \dots, z_S(t_n)]^T$, \mathcal{N} denotes a normal density, σ^2 the noise variance, \mathbf{I} the identity matrix and, as above, \mathbf{V} is the vector of maximum reaction rates. The appropriateness of normality, additivity and the uncorrelatedness of errors necessarily depends on the data-generating and measurement processes, as well as the time intervals $t_j - t_{j-1}$ between consecutive observations, as discussed in Oates and Mukherjee (2012). This approximation has the crucial advantage of rendering the local reaction graphs G_S statistically orthogonal, such that each may be estimated independently (see Hill *et al.*, 2012). Iterating over $S \in \mathcal{V}$ permits inference concerning the complete reaction graph G .

2.4.2 Bayesian Inference CheMA 1.0 uses truncated normal priors $\mathcal{N}_T(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with parameters $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ inherited from the corresponding untruncated distribution. Truncation ensures non-negativity of parameters, whilst normality facilitates partial conjugacy (see below); additional information on truncated normals is provided in the Supplementary Information. To simplify notation, we consider a specific variable S and candidate model G_S and omit the subscript in what follows. In order to elicit hyperparameters $\boldsymbol{\mu}, \boldsymbol{\Sigma}$, we follow Xu *et al.* (2010) and assume all processes occur on observable time and concentration scales, that is $\boldsymbol{\mu}_V, \boldsymbol{\mu}_K \sim \mathcal{O}(1)$, reflecting that the data are normalized *a priori*. For prior covariance of Michaelis-Menten parameters $\boldsymbol{\Sigma}_K$ we assume independence of the components K_i , so that $p(\mathbf{K}) = \mathcal{N}_T(\mathbf{K}; \boldsymbol{\mu}_K, \nu \mathbf{I})$, where $\boldsymbol{\mu}_K, \nu$ are hyperparameters. For the prior covariance $\boldsymbol{\Sigma}_V$ of maximum reaction rates we take a unit information formulation of the truncated g -prior, so that $p(\mathbf{V} | \mathbf{K}, \sigma) = \mathcal{N}_T(\mathbf{V}; \boldsymbol{\mu}_V, n\sigma^2(\mathbf{D}'\mathbf{D})^{-1})$ and for the noise parameter we use a Jeffreys prior $p(\sigma) \propto 1/\sigma$. These latter choices render the formulation partially conjugate, with the conditional density $p(\mathbf{V}, \sigma | \mathbf{K}, \mathcal{D})$ given in closed form as

$$p(\mathbf{V}, \sigma | \mathbf{K}, \mathcal{D}) = \mathcal{N}_T(\mathbf{V}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \mathcal{IG}(\sigma; a, b), \quad (7)$$

where $\boldsymbol{\mu} = \mathbf{1}/(n+1) + n/(n+1) \times (\mathbf{D}'\mathbf{D})^{-1} \mathbf{D}'\mathbf{z}$, $\boldsymbol{\Sigma} = \sigma^2 n/(n+1) \times (\mathbf{D}'\mathbf{D})^{-1}$, $a = (n-1)/2$, $b = (1/2)(\mathbf{1}'\mathbf{D}'\mathbf{D}\mathbf{1}/n + \mathbf{z}'\mathbf{z} - n/(n+1) \times \mathbf{z}'\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1} \mathbf{D}'\mathbf{z})$ and $\mathcal{IG}(\bullet; a, b)$ is an inverse gamma density with shape and scale parameters a, b respectively.

2.4.3 Marginal Likelihood Partial conjugacy of CheMA 1.0 permits an efficient Metropolis-within-Gibbs Markov chain Monte Carlo (MCMC) sampling scheme for the parameter posterior distribution. The conditional density $p(\mathbf{V}, \sigma | \mathbf{K}, G_S, \mathcal{D})$ is given in closed form as in Eqn. 7 above, while a Metropolis-Hastings acceptance step allows sampling from the remaining conditional $p(\mathbf{K} | \mathbf{V}, \sigma, G_S, \mathcal{D})$. To estimate marginal likelihoods from sampler output we exploit partial conjugacy and use the method of Chib

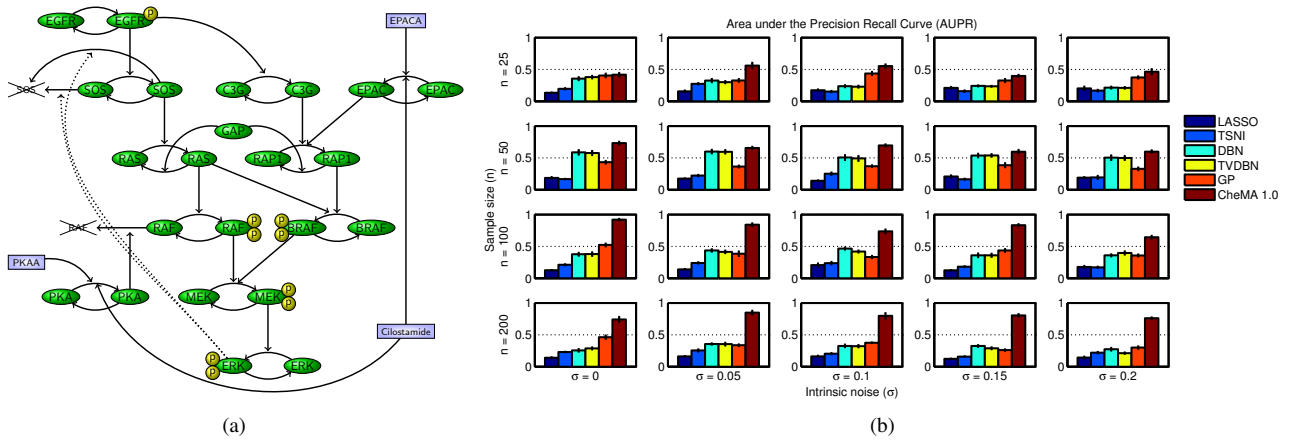


Fig. 2: Network inference, simulation study. (a) Reaction graph G for the MAPK signaling pathway, due to Xu *et al.* (2010). [The model, based on enzyme kinetics, uses Michaelis-Menten equations to capture a variety of post-translational modifications including phosphorylation.] (b) Area under the precision-recall curve (AUPR; with respect to the true causal network $N(G)$) for varying sample size n and noise level σ . [Network inference methods: (i) LASSO, ℓ_1 -penalized regression, (ii) TSNI, ℓ_2 -penalized regression, (iii) DBN, dynamic Bayesian networks, (iv) TVDBN, time-varying DBNs, (v) GP, nonparametric regression, (vi) CheMA 1.0, based on chemical kinetic models. Error bars display standard error computed over 5 independent datasets. Full details provided in Supplementary Information.]

and Jeliaskov (2001), evaluating the identity

$$p(\mathcal{D}|G_S) = \frac{p(\mathcal{D}|\mathbf{V}, \mathbf{K}, \sigma, G_S)p(\mathbf{V}, \mathbf{K}, \sigma|G_S)}{p(\mathbf{V}, \mathbf{K}, \sigma|\mathcal{D}, G_S)} \quad (8)$$

at particular parameter values $\mathbf{V}, \mathbf{K}, \sigma$ using a Monte Carlo estimate of the posterior ordinate $p(\mathbf{V}, \mathbf{K}, \sigma|\mathcal{D}, G_S)$. Since inference in CheMA 1.0 decomposes over proteins $X_i \in \mathcal{V}$ and for a given protein, over local models G_i , the computations were parallelized (full details and software provided as Supplement). Alternatively MCMC could be employed over the discrete space of reaction graphs (Ellis and Wong, 2008) or the joint space of graphs and parameters (Oates *et al.*, 2012).

2.4.4 Interventions on the System In interventional experiments, data are obtained under treatments that externally influence network edges or nodes. Inhibitors of protein phosphorylation are now increasingly available; such inhibitors typically bind to the kinase domain of their target, preventing enzymatic activity. We consider such inhibitors in biological experiments below. Within CheMA 1.0 we model inhibition by setting to zero those terms in the design matrix $\mathbf{D}_{G,S}$ corresponding to the inhibited enzyme E in the treated samples (“perfect certain” interventions in the terminology of Eaton and Murphy, 2007; Spencer *et al.*, 2012). This removes the causal influence of E for the inhibited samples.

3 RESULTS

3.1 Hyperparameter Specification and Sensitivity

For CheMA 1.0 we set hyperparameters $\mu_V = \mu_K = 1$, $\nu = 1/2$, and the maximum in-degree constraint $c_1 = 2$; we investigated sensitivity by varying these parameters within (a) a toy model of signaling (SFig. 3a-c) and (b) in a subset of the simulations reported below (SFig. 2). Since the action of inhibition is second-order in the Taylor expansion sense, inference for inhibitor variables $\mathcal{I}_{S,E}$ may be expected to require substantially more data. Indeed, “weak identifiability” of second order terms in this context was also reported in Calderhead and Girolami (2011). A preliminary investigation based on a toy model of signaling revealed that

inference for inhibitor sets $\mathcal{I}_{S,E}$ from typical sample sizes was extremely challenging (SFig. 3(d)). Combined with computational considerations, we decided to fix $c_2 = 0$ for subsequent experiments; that is, we did not include inhibitory regulation in the reaction graph. Further diagnostics, including MCMC convergence, are presented in the Supplementary Information.

3.2 In Silico MAPK Pathway

Data were generated from a mechanistic model of the MAPK signaling pathway due to Xu *et al.* (2010), specified by a system of 25 ODEs of Michaelis-Menten type whose reaction graph is shown in Fig. 2a. This archetypal protein signaling system provides an ideal test bed, since the causal graph is known and the model has been validated against experimentally obtained data (Xu *et al.*, 2010). Following Oates and Mukherjee (2012) the Xu *et al.* model was transformed into an SDE with intrinsic noise σ . Full details of the simulation set-up appear in Supplementary Information.

For inference of the network $N(G)$, we compared our approach to existing network inference methods that are compatible with time course data: (i) ℓ_1 -penalized regression (“LASSO”), (ii) Time Series Network Identification (“TSNI”; Bansal *et al.*, 2006, ; this is based on ℓ_2 -penalized regression), (iii) dynamic Bayesian networks (“DBN”; Hill *et al.*, 2012); (iv) time-varying DBNs (“TVDBN”; Dondelinger *et al.*, 2012) and (v) nonparametric (Gaussian process) regression with model averaging (“GP”; Åijö and Lähdesmäki, 2010). Approaches (i-iii) are based on linear difference equations, (iv) relaxes the linear assumption in a piecewise fashion, whereas (v) is a semiparametric variable selection technique. We note that since TSNI cannot deal with multiple time courses we adapted it for use in this setting. Implementation details for all methods may be found in the Supplementary Information.

To systematically assess estimation of network structure we computed the average area under the precision-recall (AUPR) and

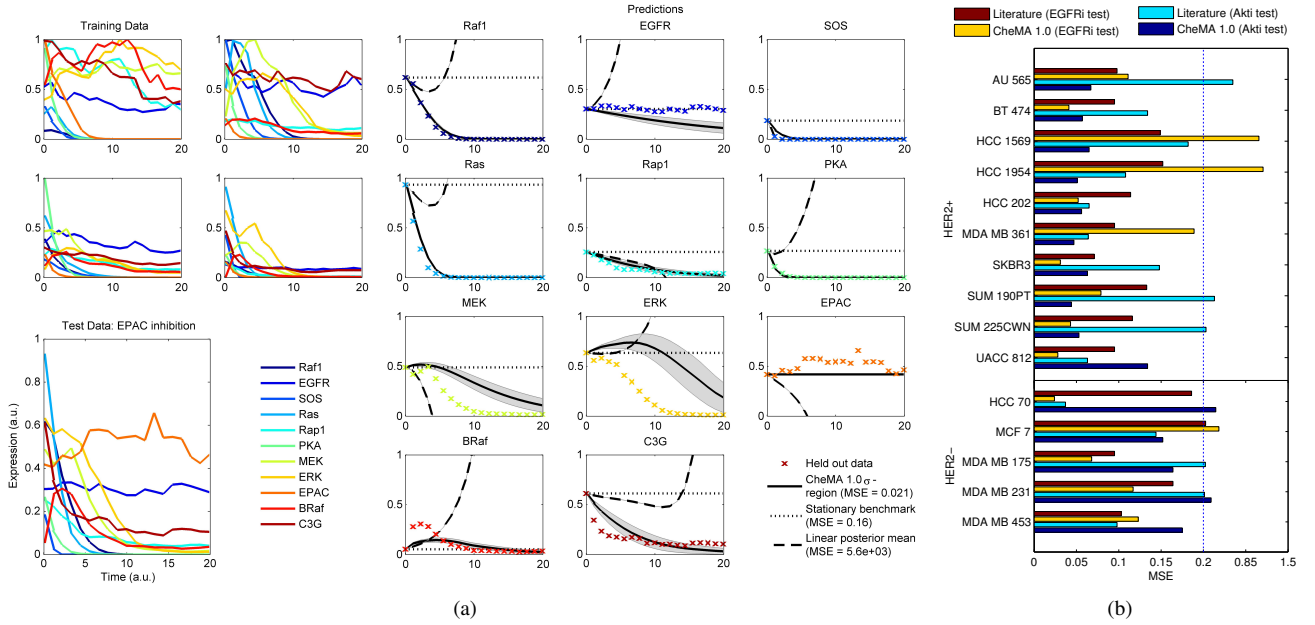


Fig. 3: Predicting dynamical response to a novel intervention: (a) Predicting the effect of EPAC inhibition under the data generating model of Xu *et al.* (2010). [CheMA (solid) regions correspond to standard deviation of the posterior predictive distribution. Linear (dashed) replaces the nonlinear chemical kinetic models with simple linear models. The stationary benchmark (dotted) simply uses the initial data point as an estimate for all later data points. The true test data are displayed as crosses. Here $n = 100$, $\sigma = 0.1$.] (b) Assessing prediction over a panel of 15 breast cancer cell lines. [Training data were time series under treatment with a single inhibitor; test data represented a second, held-out inhibitor. Normalized mean squared error (MSE) was averaged over all protein species and all time points.]

receiver operating characteristic (AUROC) curves. Fig. 2b shows mean AUPR for all approaches, for 20 regimes of sample size n and noise σ . CheMA 1.0 performs consistently well in all regimes, and outperforms (i-v) substantially at the larger sample sizes. It is interesting to note that the linear and piecewise linear DBNs (iii-iv) perform better at moderate sample sizes compared to higher sample sizes, possibly due to inconsistency arising from model misspecification. AUROC results (SFig. 6) showed a broadly similar pattern, with CheMA 1.0 offering gains at larger sample sizes. For the kinetic parameters, however, we found that CheMA 1.0 struggled to precisely recover the true values $\theta = \{V, K\}$, even when the reaction graph G was known (SFig. 8). The posterior distribution over rate constants V was much more informative than the posterior distribution over Michaelis-Menten parameters K , consistent with the “weak identifiability” of kinase inhibitors that we found in Sec. 3.1.

To investigate dynamical prediction in the setting where neither reaction graph nor parameters are known, we generated data from an unseen intervention and assessed ability to predict the resulting dynamics (details of the simulation are included in the Supplement). To fix a length scale, both true and predicted trajectories were normalized by maximum protein expression in the test data. The quality of a predicted trajectory was then measured by the mean squared error (MSE) relative to the (held out) data points. The network inference approaches (i-v) above cannot be directly applied for prediction in this setting (although they could

in principle be adapted to do so). We therefore compared CheMA 1.0 with the analogous linear formulation, that replaces Eqn. 2 by $f_{G,S}(\mathbf{X}, \theta_S) = \beta_0 + \sum_{E \in \mathcal{E}_S} \beta_E [X_E^*]$ (see Supplementary Information for details), along with a simple, baseline estimator (the “stationary benchmark”) that presumes protein concentrations do not change with time. Fig. 3a displays predictions that result from EPAC inhibition. Here CheMA 1.0 provides qualitatively correct prediction, whereas the linear analogue rapidly diverges to infinity (due to poorly estimated eigenvalues). We therefore focused only on short term prediction, specifically the first 25% of the time course, for which linear models may yet prove useful. Over all simulation regimes and experiments, including at small sample sizes, we found that our approach produced significantly lower MSE than both the linear and benchmark models ($\text{MSE}_{\text{CheMA 1.0}} = 0.061$, $\text{MSE}_{\text{Lin.}} = 2.55$, $\text{MSE}_{\text{Bench.}} = 0.199$). Furthermore CheMA 1.0 consistently produced lowest MSE at all fixed values of n and σ (SFig. 10; $p < 0.001$ binomial test).

3.3 In Vitro Signaling

Next, we considered experimental data obtained using reverse-phase protein arrays (Hennessy *et al.*, 2010) from 15 human breast cancer cell lines, of which 10 were of HER2+ subtype (Neve *et al.*, 2006). These data comprised observations for key phosphoproteins Akt, EGFR, MEK, GSK3ab, S6, 4EBP1 and their unphosphorylated counterparts. Data were acquired under pretreatment with inhibitors Lapatinib (“EGFRi”; an EGFR/HER2 inhibitor), GSK690693

(“Akti”; an Akt inhibitor) and without inhibition (DMSO) at 0.5, 1, 2, 4 and 8 hours following serum stimulation, giving a total of $n = 15$ observations of each species in each cell line (see Supplementary Information for full experimental protocol).

At present, inferred network topologies for the cell lines cannot be rigorously assessed since the true cell line-specific networks are not known. Inferred topologies partially concord with known signaling (SFig. 11), but the latter is based mainly on studies using wild type cells and may not reflect networks in genetically perturbed cancer lines. Therefore, for an unbiased test, we considered the problem of prediction of trajectories under an unseen intervention. We sought to compare performance of CheMA 1.0 against a literature-based ODE model (reaction graph G fixed according to literature and dynamics f_G as described above) fitted to training data. No prior information concerning specific chemical reactions was provided to CheMA 1.0. This problem is highly nontrivial due to the small sample size, uneven sampling times and the complex observation process associated with proteomic assay data.

Training on DMSO and EGFRi (or AKTi) data, we assessed ability to predict the full dynamic response to Akt (or EGFR) inhibition. In this way, each held-out test set contained trajectories under a completely unseen intervention. By considering all 15 cell lines, giving 30 held-out datasets, we found that in 19 out of 30 prediction problems CheMA 1.0 outperformed the literature predictor (Fig. 3b). As for the simulated data, the linear model was not well-behaved for prediction (SFig. 12) and is not shown. In the Akti test, of the 10 HER2+ cell lines 9 were better predicted by CheMA 1.0 compared to literature prediction ($p = 0.01$, binomial test; $\text{MSE}_{\text{CheMA 1.0}} = 0.064$ vs $\text{MSE}_{\text{Lit.}} = 0.274$). Conversely 4 out of 5 HER2- lines were better predicted by literature ($\text{MSE}_{\text{Lit.}} = 0.145$ vs $\text{MSE}_{\text{CheMA 1.0}} = 0.240$), suggesting that signaling network topology in HER2+ lines may differ to the (wild-type) literature topology. This agrees with biological understanding of HER2+ cell lines (Neve *et al.*, 2006) and is encouraging from the perspective of CheMA, since *a priori* it is far from clear whether the training data, which involved only $p = 6$ species and $n = 10$ data points, contain sufficient information to predict the effect of an unseen intervention, even approximately. However, in two of the failure cases (HCC 1569, HCC 1954; EGFRi test) CheMA 1.0 produced extremely poor predictions ($\text{MSE}_{\text{CheMA 1.0}} > 1$), likely due to the small training sample size.

4 DISCUSSION

We proposed a general framework for using chemical kinetics for network inference and dynamical prediction. The use of chemical kinetics can be expected to contribute gains in causal inference since the underlying models are not structurally symmetric, allowing causal directionality to be established (Peters *et al.*, 2011). In empirical results we found that whilst CheMA 1.0 struggled to identify kinetic parameters from data, it was nevertheless able to identify the causal network; this discrepancy is explained by the fact that the latter is in a sense a projection of the former, and can be identifiable even when the full set of parameters are not.

An important challenge in systems biology is to predict the effect on signaling of a novel intervention, such as a drug treatment. At present dynamical predictions in systems biology require a known chemical reaction graph, for instance taken from literature; a system

of ODEs is usually specified based on such a graph and used for prediction. However in many settings, the chemical reaction graph may differ depending on (e.g.) cell type or disease state and cannot be assumed known. In contrast, CheMA shows how prediction of dynamical behavior may be possible even when the reaction graph itself is entirely unknown *a priori*. Unlike more convenient linear or discrete formulations (e.g. Maathuis *et al.*, 2009), our use of chemical kinetic models provides interpretable predictions. For example the dynamic behavior of phosphoprotein concentrations obtained under our methodology are physically plausible (i.e. smooth, bounded and non-negative). Furthermore, by averaging predictions over reaction graphs, our approach should provide robustness in (typical) situations where it is unreasonable to expect to identify G precisely.

Several improvements can be made to the CheMA 1.0 implementation reported here, of which we highlight two: (i) Gradient matching (rather than numerical solution of the automatically-generated dynamical systems) can help to relieve the computational demands associated with exploration the large model spaces, but the Euler approximations we used for this purpose are crude. Improved gradient matching should be possible (at the expense of requiring more time points) via higher-order expansions, or (at the expense of additional modeling assumptions) kernel regression, the penalized likelihood approaches of Ramsay *et al.* (2007); González *et al.* (2013), or the Bayesian approach of Dondelinger *et al.* (2013). (ii) CheMA 1.0 does not explicitly distinguish between process noise and observation noise; an interesting direction for further research would be to incorporate an explicit observation model.

Two ongoing challenges in Bayesian computation relevant to CheMA include inference of model parameters and computation of marginal likelihoods for model selection. The first has been tackled from many directions, including approximate Bayesian computation (Toni *et al.*, 2009), Gaussian processes approximations (Dondelinger *et al.*, 2013), MCMC (Wilkinson, 2010), and particle filtering (Quach *et al.*, 2007). The second question is a comparatively under-developed area of statistical research, with candidate approaches including variational approximations (Rue *et al.*, 2009) and MCMC (Vyshemirsky and Girolami, 2008). In general the computational burden of CheMA will be higher than many methods (see Supplementary Information). By way of demonstration, Bayesian inference and prediction for a system of 27 protein species required over 12 hours (serial) computational time. In contrast, linear or discrete models offer improved scalability to high-dimensional systems by permitting closed form expression of model selection criteria. Thus, CheMA can complement existing methodologies but is not at present applicable to truly high-dimensional problems with hundreds or thousands of nodes.

Finally we note the following caveats: (i) The automatic generation of kinetic equations limits the extent to which in-depth knowledge about particular biochemical processes and dynamics may be incorporated. (ii) Our empirical results suggest that more complex interactions, including kinase inhibition, can be extremely difficult to identify from time series data. (iii) The form of kinetics used here will likely be sub-optimal when the assumptions of the Michaelis-Menten approximation are violated. (iv) Larger training and test datasets may be needed to allow truly effective trajectory prediction and comprehensive assessment of performance.

ACKNOWLEDGEMENT

This work was supported by the US Department of Energy [DE-AC02-05CH11231]; US National Institute of Health, National Cancer Institute [U54 CA 112970, P50 CA 58207]; UK Engineering and Physical Sciences Research Council [EP/E501311/1]; and Netherlands Organisation for Scientific Research [Cancer Systems Biology Center].

REFERENCES

- Äijö, T., and Lähdesmäki, H. (2010) Learning Gene Regulatory Networks From Gene Expression Measurements Using Non-Parametric Molecular Kinetics. *Bioinformatics*, 25, 2937-2944.
- Bansal, M. *et al.* (2006) Inference of Gene Regulatory Networks and Compound Mode of Action From Time Course Gene Expression Profiles. *Bioinformatics*, 22, 815-822.
- Calderhead, B., and Girolami, M. (2011) Statistical Analysis of Nonlinear Dynamical Systems Using Differential Geometric Sampling Methods. *J R Soc Interface Focus*, 1, 821-835.
- Chen, W. W. *et al.* (2009) Input-Output Behavior of ErbB Signaling Pathways as Revealed by a Mass Action Model Trained Against Dynamic Data. *Mol Syst Biol*, 5, 239.
- Chib, S., and Jeliazkov, I. (2001). Marginal Likelihood From the Metropolis-Hastings Output. *J Am Stat Assoc*, 96(453), 270-281.
- Craciun, G., and Pantea, C. (2008) Identifiability of Chemical Reaction Networks. *J Math Chem*, 44(1), 244-259.
- Dondelinger, F., Lèbre, S., and Husmeier, D. (2012) Non-Homogeneous Dynamic Bayesian Networks With Bayesian Regularization for Inferring Gene Regulatory Networks With Gradually Time-Varying Structure. *Mach Learn*, 90, 191-230.
- Dondelinger, F. *et al.* (2013) ODE parameter inference using adaptive gradient matching with Gaussian processes. *In: Sixteenth International Conference on Artificial Intelligence and Statistics*, 216-228.
- Eaton, D., and Murphy, K. (2007) Exact Bayesian Structure Learning From Uncertain Interventions. *In: Eleventh International Conference on Artificial Intelligence and Statistics*, 2, 107-114.
- Ellis, B., and Wong, W. (2008). Learning Causal Bayesian Network Structures From Experimental Data. *J Am Stat Assoc*, 103(482), 778-789.
- Gardner, T. *et al.* (2003) Inferring Genetic Networks and Identifying Compound Mode of Action via Expression Profiling. *Science*, 301(5629), 102-105.
- González, J. *et al.* (2013) Inferring Latent Gene Regulatory Network Kinetics. *Stat Appl Genet Mol*, 12(1), 109-127.
- Hennessy, B.T. *et al.* (2010) A Technical Assessment of the Utility of Reverse Phase Protein Arrays for the Study of the Functional Proteome in Nonmicrodissected Human Breast Cancer. *Clinical Proteomics*, 6(4), 129-151.
- Hill, S. M. *et al.* (2012) Bayesian Inference of Signaling Network Topology in a Cancer Cell Line. *Bioinformatics*, 28, 2804-2810.
- Honkela, A. *et al.* (2010) Model-Based Method for Transcription Factor Target Identification With Limited Data. *Proc Natl Acad Sci USA*, 107, 7793-7798.
- Kholodenko, B. (2006). Cell-Signalling Dynamics in Time and Space. *Nat Rev Mol Cell Bio*, 7(3):165-176.
- Leskovac, V. (2003) Comprehensive Enzyme Kinetics. Kluwer Academic/Plenum Publishers, New York.
- Maathuis, M. H. *et al.* (2009) Estimating High-Dimensional Intervention Effects From Observational Data. *Ann Stat*, 37, 3133-3164.
- Mukherjee, S., and Speed, T. (2008) Network Inference Using Informative Priors. *Proc Natl Acad Sci USA*, 105, 14313-14318.
- Nachman, I. *et al.* (2004) Inferring Quantitative Models of Regulatory Networks From Expression Data. *Bioinformatics*, 20(Suppl 1), i248-i256.
- Nelander, S. *et al.* (2008) Models From Experiments: Combinatorial Drug Perturbations of Cancer Cells. *Mol Syst Biol*, 4, 216.
- Neve, R. *et al.* (2006) A Collection of Breast Cancer Cell Lines for the Study of Functionally Distinct Cancer Subtypes. *Cancer Cell*, 10, 515-527.
- Oates, C. J., and Mukherjee, S. (2012) Network Inference and Biological Dynamics. *Ann Appl Stat*, 6, 1209-1235.
- Oates, C. J. *et al.* (2012) Network Inference Using Steady State Data and Goldbeter-Koshland Kinetics. *Bioinformatics*, 28, 2342-2348.
- Pearl, J. (2009) Causal Inference in Statistics: An Overview. *Statistics Surveys*, 3, 96-146.
- Peters, J. *et al.* (2011) Identifiability of Causal Graphs using Functional Models. *Uncertainty in Artificial Intelligence, Proceedings of the 27th Conference*, 589-598.
- Quach, M. *et al.* (2007) Estimating Parameters and Hidden Variables in Non-Linear State-Space Models Based on ODEs for Biological Networks Inference. *Bioinformatics*, 23, 3209-3216.
- Ramsay, J.O. (2007) Parameter Estimation for Differential Equations: A Generalized Smoothing Approach. *J R Statist Soc B*, 69(5), 741-796.
- Rue, H. *et al.* (2009) Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations. *J R Statist Soc B*, 71, 319-392.
- Sachs, K. *et al.* (2005) Causal Protein-Signaling Networks Derived From Multiparameter Single-Cell Data. *Science*, 308(5721), 523-529.
- Scott, J. G., and Berger, J. O. (2010) Bayes and Empirical-Bayes Multiplicity Adjustment in the Variable-Selection Problem. *Ann Stat*, 38, 2587-2619.
- Spencer, S.E.F., Hill, S.M., Mukherjee, S. (2012) Dynamic Bayesian Networks for Interventional Data. *CRISM Working Paper, University of Warwick* 12:24.
- Toni, T. *et al.* (2009) Approximate Bayesian Computation Scheme for Parameter Inference and Model Selection in Dynamical Systems. *J R Soc Interface*, 6(31), 187-202.
- Vysheirsky, V., and Girolami, M. (2008) Bayesian Ranking of Biochemical System Models. *Bioinformatics*, 24, 833-839.
- Werhli, A. *et al.* (2006) Comparative Evaluation of Reverse Engineering Gene Regulatory Networks With Relevance Networks, Graphical Gaussian Models and Bayesian Networks. *Bioinformatics*, 22, 2523-2531.
- Wilkinson, D. (2010) Parameter Inference for Stochastic Kinetic Models of Bacterial Gene Regulation: A Bayesian Approach to Systems Biology. *Bayesian Statistics*, 9, 679-706.
- Xu, T.-R. *et al.* (2010) Inferring Signaling Pathway Topologies From Multiple Perturbation Measurements of Specific Biochemical Species. *Science Signalling*, 3(113), ra20.